

Proteomics Primer

Advancing Proteomics for the Early Diagnosis and Treatment of Cancer

The National Cancer Institute (NCI) recognizes the promise of clinical proteomics for the early detection and treatment of cancer. For this reason, NCI has taken the lead in the scientific community to accelerate the application of proteomic technologies to the clinical setting as part of their efforts to eliminate cancer.

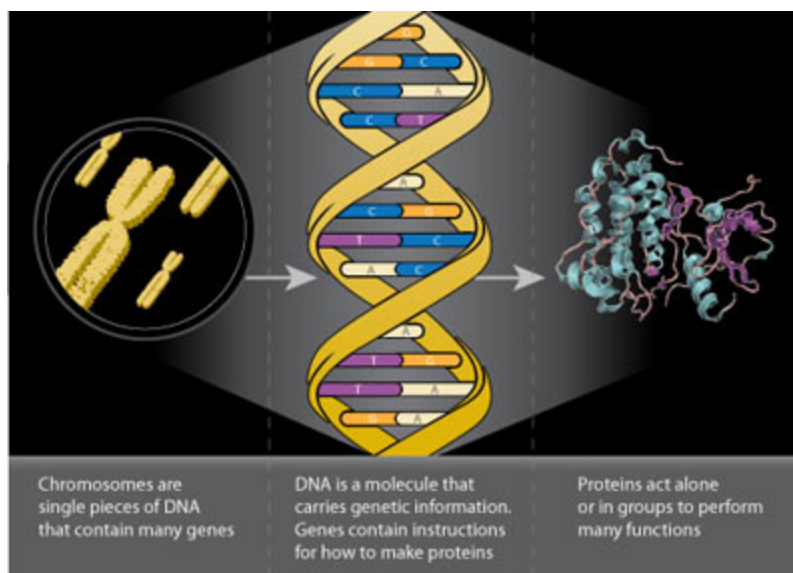
The NCI is funding \$104 million over five years to address this challenge through the Cancer Proteomic Technologies for Cancer (CPTC) initiative. This initiative is bringing together the best minds in the field of proteomics (Proteomic Technology Specialists, Clinical Cancer Researchers, Bioinformaticians, Biostatisticians, and Biologists) to ultimately lay the foundation for the next generation of molecular (proteomic) diagnostics.

What is Proteomics?

The term “**proteome**” refers to all of the proteins in a cell, tissue, or organism, while “**clinical proteomics**” refers to the study of proteomes in health and disease.

Why Proteomics?

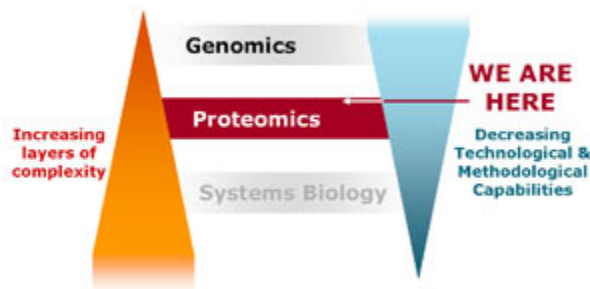
The greatest promise for the early detection and treatment of cancer lies in the ability to find valid molecular indicators, or biomarkers, of the disease. Progress in cancer genetics has been rapid, but this only provides us with a glimpse of what *may* occur: we need to measure what is happening in a patient in real time, and that means finding tell-tale protein biomarkers. This is because genes are only the “recipes” of the cell: the proteins encoded by the genes are ultimately the critical molecular players that drive both normal and disease physiology.



How Does the Proteome Compare to the Genome?

The biggest conceptual challenge inherent in proteomics lies in the proteome's increased degree of complexity compared to the genome. For example:

- **One gene can encode more than one protein** (even up to 1,000). The human genome contains about 21,000 protein-encoding genes, but the total number of proteins in human cells is estimated to be between 250,000 to one million.
- **Proteins are continually moving and undergoing changes** such as binding to a cell membrane, partnering with other proteins, or breaking into two or more pieces. The genome, on the other hand, is relatively static.
- **Cells are continually modifying proteins** once they are produced. As a result, the types of proteins measured can vary considerably from one person to another, under different environmental conditions, or even within the same person at different ages or states of health.
- **Proteins exist in a wide range of concentrations in the body.** For example, the concentration of the protein albumin in blood is more than a billion times greater than that of interleukin-6, making it extremely difficult to find the low abundance proteins in a mixture. Scientists believe that the most important proteins for cancer may be those found in the smallest concentrations.



How Can Proteomics Help Diagnose and Treat Cancer?

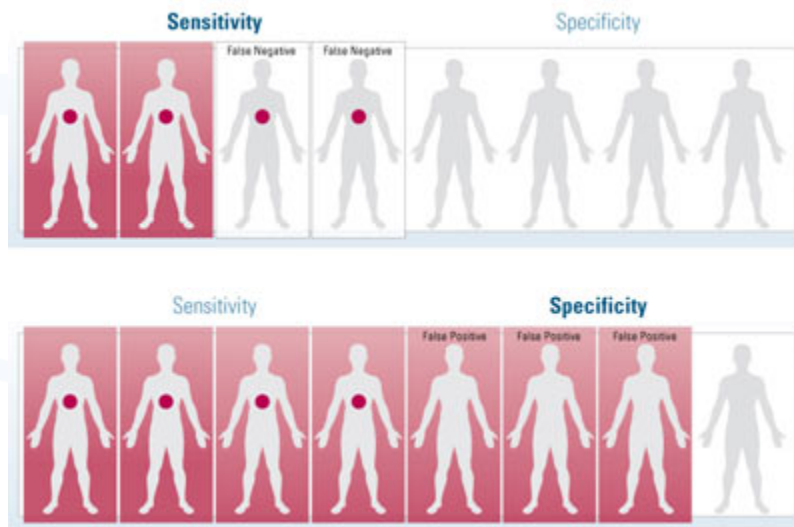
Of critical importance to cancer research was the finding that tumors “leak” proteins into blood, urine, and other accessible bodily fluids. This insight has led to the possibility of diagnosing cancer at an early stage simply by collecting such fluids and testing them for the presence of cancer-related molecules, or “biomarkers.”

The earlier a patient’s cancer is diagnosed the more treatable it is by surgery, radiation or chemotherapy. Biomarkers found in blood and other fluids might also be valuable for monitoring the response to cancer during treatment or detecting the recurrence of tumors after treatment.

Certain blood proteins are already being used as cancer biomarkers. For example, elevated levels of prostate specific antigen (PSA) suggest the presence of prostate cancer, while elevated levels of cancer antigen 125 (CA-125) suggest recurrent ovarian cancer. Unfortunately, both tests may result in “false negatives” – failing to detect cancer in those who have it (poor sensitivity), or “false positives” – testing positive for the presence of cancer in people who are actually cancer-free (poor specificity).

Scientists have proposed that in order to develop more sensitive and specific cancer diagnostic tests, many biomarkers should be measured simultaneously. It is thought that patterns revealed in a panel of biomarker proteins associated with a form of cancer – known as a “protein signature” – might have better diagnostic and predictive capabilities than the current single-marker approach.

The good news is that there are well over 1,000 cancer protein biomarker candidates that have been described in the scientific literature over the past decade, and this list continues to grow. The sobering news is that very few of these candidates have made it through the clinic.



What's the Problem?

The proteomics community is struggling with this discrepancy. The NCI and experts in the field believe that the problem is due to a complete lack of standards in the technologies and methods that are used to discover protein biomarkers. This helps to explain why many of the biomarker candidates turn out to be merely artifacts and fail to be clinically validated. This challenge is limiting the number of cancer biomarker tests that are available to the public.

The research community requires improved proteomic technologies in order to support the development of diagnostic tests that can be applied in a clinical setting. To achieve these objectives, several advances are required:

- **Optimization of proteomic technologies and development of appropriate standards.** Current and emerging protein measurement technologies must be optimized and calibrated through the use of standards to produce comparable results between different laboratories. If data are not reproducible, then the field will fail.
- **Availability of high-quality reagents.** In particular, capture reagents that can be used in protein microarrays, as well as other techniques used to measure proteins, are essential resources.

- **Technologies that can quantify proteins across the entire concentration range as well as detect modified versions of proteins.** Immense variations in protein concentrations and type are found within cells and fluids. Current technologies are often unable to identify and analyze all of the proteins in samples with a wide range of concentrations.
- **Common bioinformatics resources, with shared algorithms and standards for processing, analyzing and storing proteomic data.** Proteomic informatics tools that permit data sharing and computation among laboratories are essential for rapid progress in the field.
- **Standardized procedures for processing and storing biological samples used in proteomics research.** Methods of biological sample preparation must be more consistent to reduce variability in experimental results. Uniform sample quality, as well as access to large numbers of high-quality samples, will lead to more reliable results.

Another development needed for the advancement of proteomics is an interdisciplinary team approach to science. No one laboratory working on its own could possibly examine all of the potential biomarkers, develop all of the necessary technologies for isolating and validating biomarkers for research or clinical use, or assemble all of the pieces of evidence required to understand the molecular mechanisms of cancer. It will require many laboratories working together to accomplish these goals.

In many ways, the challenges of developing clinical proteomic technologies and finding cancer biomarkers are comparable to those faced by **The Human Genome Project**, in which the work of sequencing the human genome was divided up among many laboratories throughout the world. ***All of the laboratories used standardized methods of collecting and analyzing their data so that the data could be assembled together at the end of the project.***

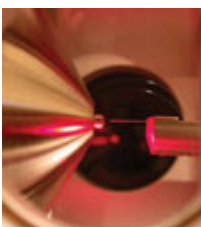
Similarly, standards are needed in cancer proteomics – standards for collecting and processing clinical samples, conducting experiments, and collecting and analyzing data – so that teams of scientists in different laboratories can collate and analyze their data to achieve meaningful results.

Proteomic Technologies

The accessibility of cancer-related proteins in bodily fluids and tissues has triggered extensive protein-focused research. But a lack of reliable methods for protein identification and measurement has led to pervasive problems with reproducibility and comparability of research results. The lack of standards and reagents have made it nearly impossible to develop, manage, interpret, and compare large quantities of proteomic data, and that in turn slows the translation of discoveries to clinical application.

Described below are some of the technologies being used that are advancing our understanding of protein biology.

Mass Spectrometry



Mass spectrometry (MS) is an evolving technology that allows scientists to detect and identify ever-smaller amounts of proteins. The method is very precise, distinguishing proteins that differ in composition by a single hydrogen atom, the smallest atom. It is also extremely fast. The entire process, from blood collection to data analysis, can take less than one minute.

Mass spectrometry, despite its potential, is not yet capable of separating the complex protein mixtures from unprocessed human biospecimens. New technologies are required to reduce the complexity of biospecimens by enriching for proteins of interest, and to enhance the range and sensitivity of the instrumentation.

Protein Microarrays



Protein microarrays are powerful tools for capturing and measuring proteins from blood and other body fluids and tissues. A protein microarray typically consists of a small piece of glass or plastic that is coated with thousands of “capture reagents” (molecules that can “grab” specific proteins). This technology allows scientists to isolate and study many potential biomarker proteins.

Because protein microarray elements can be miniaturized to contain tens of thousands of capture features arranged in a grid, each specific for a given protein, they are considered a multiplexed device – for example, they can test for multiple biomarkers simultaneously, which is excellent for clinical use.

Nanotechnologies

Nanotechnology is the creation of manufacturing devices and components that range from 1–100 nanometers. A nanometer is one billionth of a meter, or 1/80,000 the width of a human hair. Nanotechnology devices have the potential to greatly expand the capabilities of proteomics, addressing current limitations in selectively reaching a target protein *in vivo* through physical and biological barriers, detecting low abundance targets, and providing a “toolbox” to translate the discovery of protein biomarkers to novel therapeutics and diagnostic tests. Typical devices include nanoparticles used for the targeted delivery of anticancer drugs, energy-based therapeutics (including heat and radiation) and imaging contrast reagents. Nanowires and nanocantilever arrays can be used in biosensors that measure minute quantities of biomarkers in biological fluids.

For more information, see the [NCI Alliance for Nanotechnology in Cancer](#)

Bioinformatics

Major areas of focus in bioinformatics research include data modeling and database design, data interoperability and comparison, gene and protein expression analysis, structural predictions, vocabularies and ontologies, and systems biology modeling. In cancer research, the development of new tools in these areas are necessary to drive the collaborative, multidisciplinary effort required to push cancer research and discovery from the laboratory to clinical practice.

The NCI is addressing the need for an IT infrastructure to enable collection, analysis and sharing of huge amounts of data for inter-institutional studies through its cancer Biomedical Informatics Grid®, or [caBIG®](#), a voluntary network connecting individuals and institutions to enable the sharing of data and tools, creating a World Wide Web of cancer research.

Biospecimens

Cancer research has come to rely on biospecimens for the measurement of genetic and protein expression and the linkage of that information with clinical status and to disease pathways such as tumor growth, migration, metastasis, angiogenesis, and apoptosis (cell death). Since the process of

cancer diagnosis and treatment often begins with diagnostic biopsies followed by surgical resection of the tumor, there are many opportunities to collect valuable biospecimens for research.

The NCI has recognized the critical need for research access to large numbers of high-quality biospecimens annotated with clinical data. NCI is seeking to address the need by through its [Office of Biorepositories and Biospecimen Research](#).

Reagents

There is a growing need in the field of proteomics for high-quality, standardized reagents that can improve proteomic technologies' specificity and reproducibility. One widely used reagent in proteomic research is the antibody, a naturally occurring serum protein whose biological role requires high-antigen specificity. They have been useful as detection and capture reagents in proteomics.

Aptamer reagents show promise as an adjunct to antibodies. These nucleic acid-based molecules possess protein-binding specificity, similar to antibodies that make them useful as protein capture and detection reagents. Since they are nucleic acid-based, the technology for their synthesis and chemical modification is more mature than antibody production, and various mutation and selection protocols can be used to specify their binding properties. Standard proteomic reagents will be useful for many applications in cancer research and development, including:

- Reporter molecules that detect the presence of a target (or modifications to it) in a particular biological sample
- Capture molecules for purifying the target from a complex biological sample prior to identification and quantification using, for example, mass spectrometry
- Functional studies to validate the role of a potential therapeutic target prior to launching drug discovery or development efforts
- Reference materials for calibrating instruments or comparing different proteomic platform technologies